

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-249677

(43)Date of publication of application : 17.09.1999

(51)Int.Cl. G10L 3/00  
G10L 5/02

(21)Application number : 10-049161

(22)Date of filing : 02.03.1998

(71)Applicant : HITACHI LTD

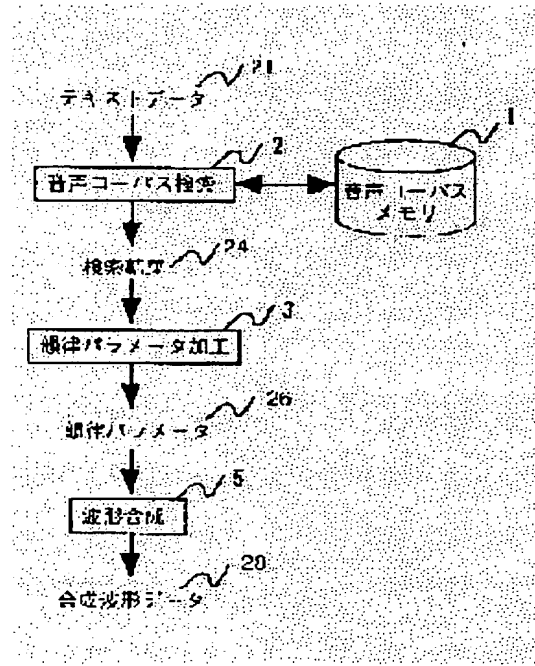
(72)Inventor : NUKAGA NOBUO  
KITAHARA YOSHINORI  
FUJITA KEIKO  
ANDOU HARU  
YAJIMA SHUNICHI

## (54) RHYTHM CONTROL METHOD FOR VOICE SYNTHESIZER

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide a text voice synthesizer synthesizing a voice with naturalness higher than a usual method by retrieving a voice corpus consisting of a sentence voice, its rhythm parameter, a morpheme/construction data, etc., extracting the data with the largest similarity with an inputted sentence, processing/correcting the rhythm parameter for the extracted data and generating the rhythm parameter.

**SOLUTION:** A sentence utterance voice is stored in a voice corpus memory 1, and a text having a rhythm near to an inputted text is retrieved, and its rhythm parameter 24 is deformed, and the rhythm parameter 26 for the inputted text is calculated. Plural rhythm parameters 26 are treated as the coordinated data, and a synthetic voice with natural intonation and rhythm and near to a natural voice is generated.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japanese Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-249677

(43) 公開日 平成11年(1999) 9月17日

(51) Int.Cl.<sup>6</sup>

識別記号

F I

G 1 0 L 3/00  
5/02

G 1 0 L 3/00  
5/02

H  
J

審査請求 未請求 請求項の数 4 O L (全 10 頁)

(21) 出願番号 特願平10-49161

(22) 出願日 平成10年(1998) 3月2日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 額賀 信尾

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(72) 発明者 北原 義典

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(72) 発明者 藤田 啓子

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(74) 代理人 弁理士 小川 勝男

最終頁に続く

(54) 【発明の名称】 音声合成装置の韻律制御方法

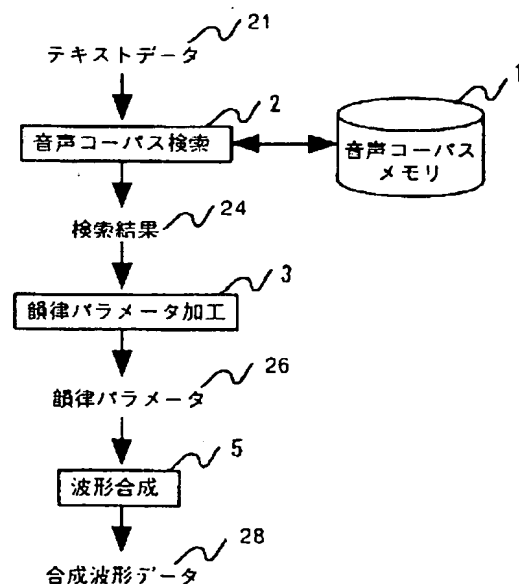
(57) 【要約】

【課題】 従来の規則音声合成方法では、各韻律パラメータを個別に扱っていたために自然性の高い音声を作成することが困難であった。

【解決手段】 文発声音声を音声コーパスメモリに記憶し、入力テキストに近い韻律を持つテキストを検索し、その韻律パラメータを変形することで、入力テキストに対する韻律パラメータを計算する。

【効果】 複数の韻律パラメータを連係したデータとして取り扱うことができるようになり、抑揚やリズムが自然で肉声に近い合成音を生成することができる。

図3



## 【特許請求の範囲】

【請求項 1】テキストデータの韻律パラメータを計算し、上記計算された韻律パラメータより合成波形を生成する音声合成装置における韻律制御方法において、上記音声合成装置は、複数のテキストデータについて、各テキストデータの人間の発声に基づく複数種の韻律パラメータを記憶する音声コーパスを備え、上記入力テキストデータと上記音声コーパスに記憶されたテキストデータとを順次比較し、上記音声コーパスより少なくともその一部が一致するテキストデータを選択し、上記選択されたテキストデータと上記入力テキストデータとの不一致部分についての上記複数種の韻律パラメータを求め、上記選択されたテキストデータと上記入力テキストデータとの一致部分についての各韻律パラメータと上記不一致部分についての各韻律パラメータとを整合させることにより、上記入力テキストデータの韻律パラメータを計算することを特徴とする韻律制御方法。

【請求項 2】請求項 1 記載の韻律制御方法において、上記複数の韻律パラメータとして、基本周波数パターンと継続時間長とを含むことを特徴とする韻律制御方法。

【請求項 3】請求項 1 記載の韻律制御方法において、上記入力テキストデータについて形態素分割を行い、各形態素  $D_i$  ( $i = 1 \sim n$ ) についてその品詞とアクセント型を求め、上記音声コーパスに記憶されたテキストデータのうち、その形態素数が一致するテキストデータの各形態素  $D'_j$  ( $j = 1 \sim n$ ) についてそれぞれ上記  $D_i$  と品詞及びアクセント型を比較することにより、上記音声コーパスに記憶されたテキストデータから少なくとも 1 つのテキストデータを選択することを特徴とする韻律制御方法。

【請求項 4】請求項 3 の韻律制御方法において、上記音声合成装置は、単語のモーラ数及びアクセント型の組み合わせに対して 1 つ以上の基本周波数パターンを記憶する単語基本周波数パターンテーブルを備え、上記選択されたテキストデータと上記入力テキストデータとの不一致部分について、上記単語基本周波数パターンテーブルを索表することにより上記不一致部分の基本周波数パターンを計算することを特徴とする韻律制御方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、テキストから音声合成するテキスト音声合成技術に関し、特に、文の抑揚や時間長を制御する韻律制御方法に関する。

## 【0002】

【従来の技術】テキスト音声合成は、入力されたテキストに対して言語処理を行い、中間的な発音記号列を生成した後、発音記号から基本周波数パターンや音素継続時間などの韻律パラメータと、同時に発音記号から各音素

の声道パラメータを求め、これらのパラメータから音声を合成する方法が一般的である。例えば、文献「音声合成技術：広瀬啓吉（音声処理技術とその応用、情報処理、pp. 984-991、1997年11月）」にはこの方式を利用した音声合成装置が記載されている。

【0003】上記の方法を利用した場合、韻律パラメータは音声の抑揚やリズム、なめらかさといった全体にかかわる自然性としての特徴を決め、声道パラメータは単語や文を構成する個々の音節の明瞭性としての特徴を決定する。

【0004】韻律パラメータのうち、基本周波数パターンのパラメータ生成方法の代表的なものとして「重畳型モデル」が挙げられる。この基本周波数パラメータの生成モデルは、句などに対応して基本周波数が緩やかに下降するフレーズ成分に、文節単位のアクセント型に対応する基本周波数の上昇・下降であるアクセント成分を重畳するものである。重畳型モデルは、人間の発声機構を模擬したモデルであるため、直感的に理解しやすく、現実の音声現象とも良く合致しているが、モデルの性能を発揮するためには高度な言語処理が必要であるという問題があった。

【0005】韻律パラメータのうち、音節（音素）継続時間長は、音節（音素）が置かれるコンテキストをはじめとする種々の要因によって変化する。継続時間長に影響を与える要因として、調音上の制約、タイミング、単語の重要度、発話区分境界の明示、発話区分内のテンポ、全体のテンポ、構文意味内容等の言語的制約がある。継続時間長制御においては、実際に観測される継続時間長データに対して、上に挙げた要因に関する影響度を統計的に分析し、その結果得られる規則を利用する方法が一般的である。そのため、多様なコンテキストにおける継続時間長モデルを構築しようとすると大規模なデータベースが必要となるという問題がある。

【0006】このような韻律パラメータの他、パワーに関するパラメータなど各種制御モデルが考案されているが、いずれのモデルも各韻律パラメータ独立のモデルであり、これら単独の制御モデルの性能向上だけでは、自然性向上に限界がある。また、文音声発声には、規則によるモデル化が困難な韻律現象があることも指摘されている。

【0007】そのため、肉声から抽出した韻律パラメータをデータベースとして記憶し、データベースを利用して韻律パラメータを生成する韻律パラメータ計算モデルが提案されている。例えば、片江ら「文型—韻律データベースを用いた定型文音声合成システム、音講論、pp. 275-276、平成8年3月」や、斎藤ら「付属語連鎖単位を音声合成単位とする規則合成音声、音講論、pp. 317-318、平成6年10月」である。しかしながら、これらは基本周波数パターンのみを韻律パラメータとするものであり、なお文音声発声の自然性

を高めるには不十分なものであった。

#### 【0008】

【発明が解決しようとする課題】本発明では、肉声の各種の韻律パラメータ（基本周波数パターン、音節継続時間長等）を相互に保存したまま編集・加工することにより、さらに自然性の高い音声を合成する音声合成装置を実現するものである。

【0009】そこで本発明では、文音声、その韻律パラメータ、形態素・構文解析データなどからなる音声コーパスを備え、音声コーパスを検索して入力文との類似度が最も大きいデータを抽出し、抽出されたデータに対して韻律パラメータの加工・修正を行うことにより韻律パラメータを生成することで、従来の方法と比較して自然性の高い音声を合成するテキスト音声合成装置を提供する。

#### 【0010】

【発明の実施の形態】図1は、本発明の音声合成装置の一実施態様の構成を示すブロック図である。6は汎用メモリ、4は韻律パラメータ計算部、5は波形合成部である。韻律パラメータ計算部4は、音声コーパスメモリ1、音声コーパス検索部2、韻律パラメータ加工部3を含む。

【0011】図2に汎用メモリ6の内容を示す。汎用メモリ6には、テキストデータ10、韻律パラメータ11、合成波形データ12、音声コーパス検索部検索結果13、及び計算処理途中のデータ14が記憶される。

【0012】図3に、本発明の音声合成処理の流れを示す。音声コーパス検索部2は、入力されたテキストデータ21の解析を行って、音声コーパスメモリ1より韻律が類似しているデータを検索する。検索結果24は韻律パラメータ加工部3に入力される。なお、音声コーパスメモリ1に、テキストデータ21と韻律が類似するデータが存在しない場合には、韻律の類似するデータが存在していないことを検索結果24として出力する。韻律パラメータ加工部3では、検索結果24とテキストデータ21から、韻律パラメータ26を計算する。波形合成部5では、韻律パラメータ26を利用して合成波形データ28を生成する。

【0013】以下、「渋谷まで渋滞しています。」というテキストデータを例として、本発明によるテキストデータから合成波形への変換方法を図を用いて説明する。

【0014】図4は、音声コーパス検索部2における処理の流れである。まず、テキストデータ31「渋谷まで渋滞しています。」を、汎用メモリ6のテキストデータ10より読み出す（ステップ101）。読み出されたテキストデータ31は、単語に分割され、単語毎に表記、読み、品詞、及びアクセント情報から構成される構造化データ列33に変換する（形態素分割処理102）。構造化データ列33は、汎用メモリ6の計算処理途中データ14として記憶する。テキストデータを単語に分割す

る方法としては、清水ら「隣接単語間の結合関係に着目したテキスト音声変換用形態素解析処理、日本音響学会誌、51巻、1号、pp. 3-13、1995」に記載の処理が利用できる。単語毎の読み及びアクセント情報は、辞書に登録されているデータを設定する。

【0015】構造化データ列について図6を参照して説明する。40は1つの単語の構造化データであり、単語の表記42、読み43、品詞44、アクセント情報45を含む。テキストデータ31は例えば、「渋谷／まで／渋滞／し／て／い／ます／。」と単語分割されるので、形態素分析結果は、各単語の構造化データの列33のとおりのこととなる。

【0016】次に、音声コーパスメモリ1から一のデータセットを読み出す（ステップ103）。図7に音声コーパスメモリ1のデータ構造を示す。音声コーパスメモリ1は、複数のデータセット401、402・・・から構成される。各データセットは、図8に示すように、文字表記データ501、文字表記データ501を発声した音声波形データ502、音声波形データ502の基本周波数パターンデータ503、音声波形データ502の継続時間長データ504から構成される。なお、データセットには、パワーなどの韻律パラメータやケプストラムなどの音響パラメータさらに、文字表記データ501の形態素分割結果を含んでも良い。

【0017】音声コーパスメモリ1のデータセットについて、図9を参照して文字表記データ601が「新宿まで運転しています。」という文であった場合を例に説明する。基本周波数パターンデータ603は、音節の始末端における基本周波数の列として記憶されている。例えば、文字表記データ601の先頭の音節「し」の始端周波数605は「214」であり、終端周波数606は「190」である。また、先頭の音節「し」の継続時間長データ604は、音節のミリ秒の単位で記憶され、子音の時間長607は「101」であり、母音の時間長608は「75」である。

【0018】次に、音声コーパス検索部2では、一のデータセット600の文字表記データ601を読み出し（ステップ104）、図10に示すように、形態素分割を行って形態素分割結果703を汎用メモリ6の計算途中のデータ14に記憶する（ステップ105）。なお、データセット600に形態素分割結果が含まれている場合には、本処理は不要である。

【0019】次に、汎用メモリ6より、テキストデータ31から求めた形態素解析結果33と、一のデータセット600の文字表記データ601の形態素分割結果703を読み出し、類似度の計算を行う（ステップ106）。

【0020】以下、図11を用いて類似度計算の例を示す。最初に、形態素解析結果33と形態素分割結果703との構造化データ数を比較する（ステップ800）。

ここでは、33と703の構造化データ数はともに「8」であり、一致する。構造化データ数が一致すれば次の判定に進み（ステップ801）、一致しなければ類似度「0」と判定し、類似度計算106の処理を終了する（ステップ802、803）。

【0021】次に構造化データ数が一致した場合には、構造化データのそれぞれについて品詞及びアクセント型の比較を行う（ステップ804）。形態素解析結果33から各構造化データ $D_i$ （ $i=1\sim n$ ）と、形態素解析結果703から各構造化データ $D'_i$ （ $i=1\sim n$ ）とを比較する。例えば、 $D_1$ 「渋谷」と $D'_1$ 「新宿」とでは、ともに品詞は「地名名詞」、アクセント型は「平板型」であり一致する。このように、全ての $D_i$ と $D'_i$ について品詞とアクセント型の比較を行って、全て一致すれば、類似度を「1」として類似度計算を終了する（ステップ808）。また、1つでも一致しない構造化データがあれば、類似度を「0」として類似度計算を終了する（ステップ807）。出力された類似度は、汎用メモリ6の計算途中のデータ14に記憶される。

【0022】次に、汎用メモリ6の計算途中のデータ14から類似度を読み出し、予め設定したしきい値（基準類似度）との比較を行い（ステップ107）、検索結果を出力する（ステップ108）。図11に記載の類似計算の場合では、予め設定したしきい値を「1」とし、類似度計算106により計算された類似度が1であれば、比較結果として「一致」を出力し、類似度が0であれば「不一致」を出力する。検索結果の出力108では、類似度比較の結果が「不一致」であれば（109）、音声コーパスメモリ1に記憶されているデータセットを順次読み出し（ステップ103）、データセットがなくなるまで類似度の計算を行う。類似度比較の結果が「一致」であれば、一致したデータセット600を汎用メモリ6の音声コーパス検索部結果13に出力する。

【0023】もし、音声コーパスメモリ1の全てのデータセットに対して上記の類似度計算処理を行った結果、基準類似度を満たすデータセットが存在しない場合には、その旨を示すデータフラグ（「類似データなしフラグ」と呼ぶ）を、汎用メモリ6の音声コーパス検索部結果13に出力する。上記の処理を経て、1つ以上の類似するデータセット、あるいは、類似データなしフラグが汎用メモリ6の音声コーパス検索部13に出力される。

【0024】図12は、韻律パラメータ加工部3における処理の流れである。まず、テキストデータ31を読み出し、発音記号列を生成する（ステップ1001）。テキストデータから発音記号列へ変換する方法は、勾坂ら「日本語単語連鎖のアクセント規則、電子情報通信学会論文誌、J66-D、No. 7、pp. 849-856、1983」を用いることができる。図13に発音記号列の例を示す。発音記号列901には、読みの情報のほかに、文節の区切れ904、文末記号905、母音の

無声化記号903、アクセント記号902などが含まれている。なお、発音記号列901の生成には、汎用メモリ6に記憶されているテキストデータ31の形態素分割結果33を利用することができる。

【0025】次に、汎用メモリ6の音声コーパス検索部検索結果13から、類似度計算処理の結果を読み出す（ステップ1002）。類似度計算処理の結果は、（1）1つ以上のデータセットである場合（1003）と、（2）類似データなしフラグである場合（1004）のいずれかである。

【0026】類似するデータセットが存在すれば、1つのデータセットを選択する。このデータセットを「選択データセット」とよぶ。テキストデータ31「渋谷まで渋滞しています」に対して、データセット600「新宿まで運転しています」（図9）が選択データセットの例となる。選択データセットは、テキストデータ31と韻律に近いデータである。なぜなら、図5及び図10に示すように両者の形態素分割結果は、「渋谷」32及び「新宿」702に対応する構造化データと、「渋滞」38及び「運転」708に対応する構造化データ以外は、全て同一の構造化データあり、また、異なる構造化データに関しても、品詞及びアクセント型が等しいからである。

【0027】そこで、テキストデータ31の韻律パラメータを計算するために、選択データセット600の韻律パラメータである基本周波数パタンデータ603及び継続時間長データ604を利用すれば、肉声の韻律パラメータと近い韻律パラメータが得られ、自然性が格段に向上する。韻律パラメータ計算の方法を、図14を用いて説明する。

【0028】まず、テキストデータの形態素分割結果1101（先の例では形態素分割結果33）と選択データセット1102（先の例では形態素分割結果701）に対して、一致部と不一致部の分離（ステップ1103）を行う。1104は分離した結果を模式的に表したものであり、構造化データ1105及び1106は一致した構造化データを、構造化データ1107及び1108は一致しなかった構造化データを示す（先の例では、構造化データ32及び702、または構造化データ38及び708である）。

【0029】次に、テキストデータ31「渋谷まで渋滞しています。」の音節数から成るデータ列を生成し、分離結果1104に基づき、一致部の韻律パラメータをコピーする（ステップ1109）。一致部の韻律パラメータは、選択データセット1102（先の例ではデータセット600）の韻律パラメータをそのまま用いる。基本周波数パタンデータ1110、継続時間長データ1111は、テキストデータ31の韻律パラメータであり、網掛けの部分は一致した構造化データに対応する音節の韻律パラメータが、空白の部分が不一致の構造化データに

対応する音節の韻律パラメータが格納されている。

【0030】次に、不一致部の音節について韻律パラメータを計算する（ステップ1112）。基本周波数パターンは、単語のモーラ数とアクセント型で1つの基本周波数パターンデータを記憶する単語基本周波数パターンテーブルを用意して、この単語基本周波数パターンテーブルから検索することで、単語基本周波数パターンを求めることができる。また、単語継続時間長は、句坂ら「規則による音声合成のための音韻時間長制御、信学論、Vol. J 67-A、No. 7、pp. 629-636、1984」を利用して求めることができる。これら公知の方法により、不一致部の単語基本周波数パターンデータ1113（1115）、不一致部の継続時間長データ1114（1116）を求めることができる。

【0031】最後に、不一致部の韻律パラメータを、一致部の韻律パラメータと滑らかに接続するように、不一致部の韻律パラメータを変形・統合する（ステップ1117）。基本周波数パターンデータに関しては、始端周波数1120及び終端周波数1121の基本周波数が、選択データセット1102（先の例ではデータセット600）における相当する基本周波数の値と一致するように、単語基本周波数パターンデータを線形に変形する。また、継続時間長に関しては、選択データセット1102における相当する単語の継続時間長をモーラ数で除した値（1モーラあたりの時間長L）を利用して、1114及び1116における1モーラあたりの時間長がLと等しくなるように、1114及び1116の継続時間長を線形伸縮する。このようにして、テキストデータ31の韻律パラメータとして、基本周波数パターンデータ1118及び継続時間長データ1119が計算される。

【0032】なお、音声コーパス検索部2から類似データなしフラグが出力された場合には、音声コーパスの韻律データを利用して韻律パラメータを計算することができない。従って、発音記号列901より、従来の方法に従って韻律パラメータを計算する（ステップ1006）。なお、この場合は音声コーパスを利用して合成した音声より自然性が劣るため、任意文を合成するために音声コーパスは大容量の記憶媒体に記憶させることが望ましく、例えば、磁気記憶媒体、光記憶媒体、光磁気記憶媒体、フラッシュメモリなどに記憶させることができ

る。また、音声コーパスと通信線を介した接続することで、複数の音声合成装置で共有することも有効である。

【0033】求められた韻律パラメータ1007は、汎用メモリ6の韻律パラメータ11に記憶する。

【0034】このように、韻律パラメータ計算部3により計算された基本周波数パターン及び該継続時間長を汎用メモリ6の韻律パラメータ11より読み出し、波形合成部5において、出力音声波形を合成する。合成された波形データは、汎用メモリ6の合成波形データ12に記憶される。

【0035】

【発明の効果】本発明により、抑揚やリズムが自然で肉声に近い合成音を生成することができる。

【図面の簡単な説明】

【図1】本発明の音声合成装置のブロック図である。

【図2】音声合成装置の汎用メモリに記憶される内容の図である。

【図3】本発明の音声合成装置における処理の流れ図である。

【図4】音声コーパス検索部における処理例である。

【図5】テキストデータから形態素分割結果への変換例である。

【図6】形態素分割結果の構造化データ列である。

【図7】音声コーパスのデータ構造例である。

【図8】音声コーパスのデータセットのデータ構造例である。

【図9】データセット例である。

【図10】データセットの文字表記データから形態素分割結果への変換例である。

【図11】類似度計算処理例である。

【図12】韻律パラメータ計算部における処理例である。

【図13】発音記号列の例である。

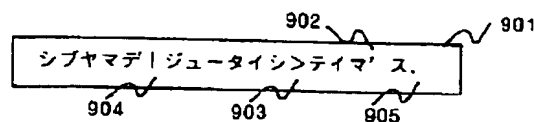
【図14】音声コーパスのデータを利用した韻律パラメータ計算例である。

【符号の説明】

1…音声コーパスメモリ、2…音声コーパス検索部、3…韻律パラメータ加工部、4…韻律パラメータ計算部、5…波形合成部、6…汎用メモリ。

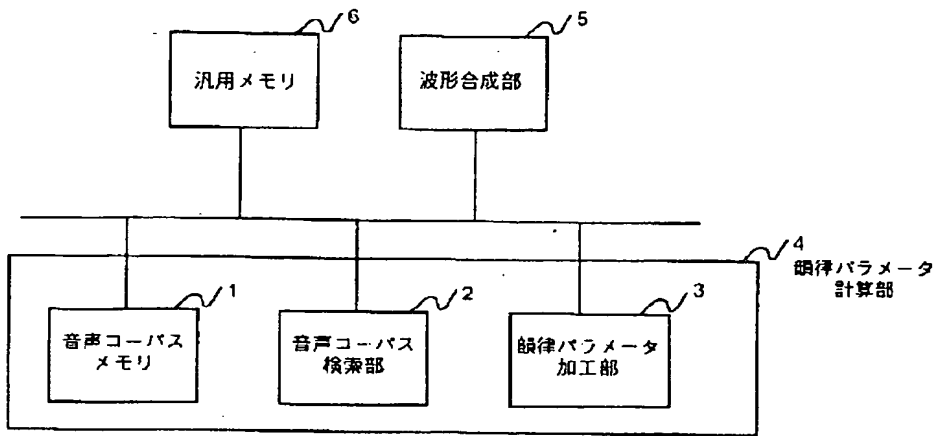
【図13】

図13



【図 1】

図 1

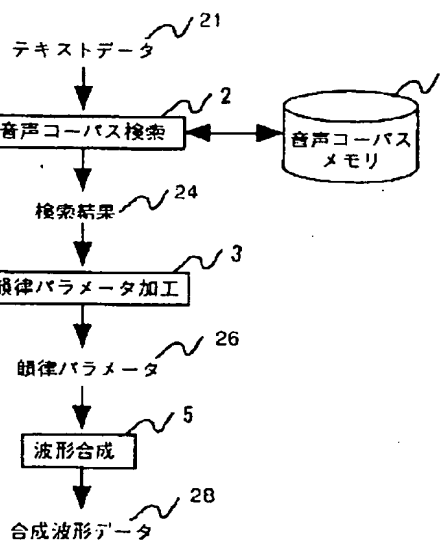
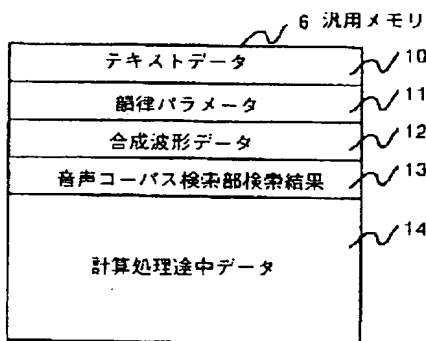


【図 2】

【図 3】

図 2

図 3

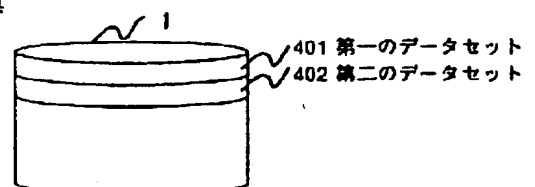
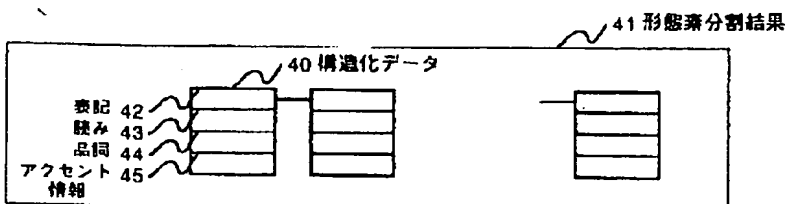


【図 6】

【図 7】

図 6

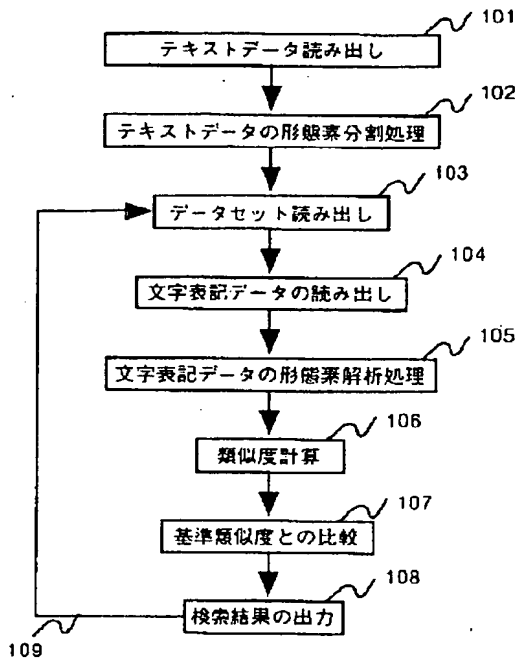
図 7





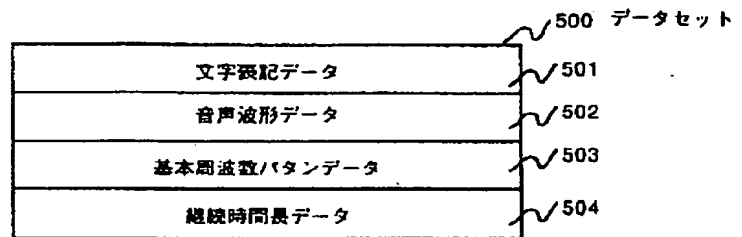
【図 4】

図 4



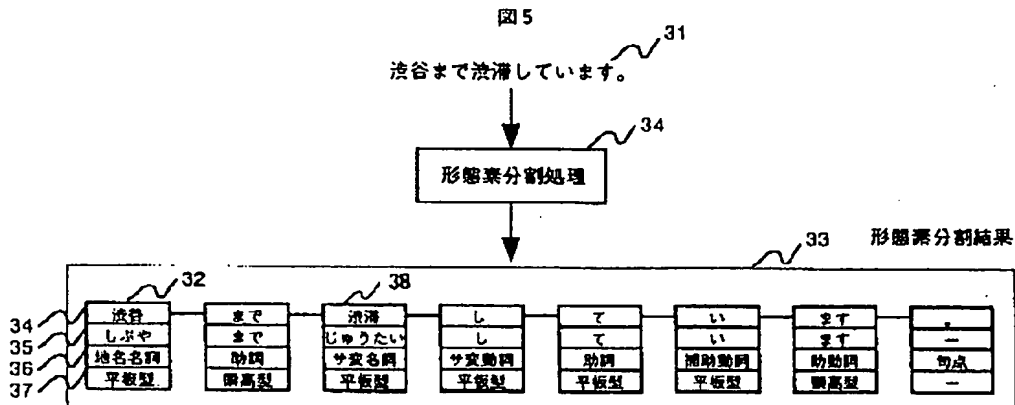
【図 8】

図 8

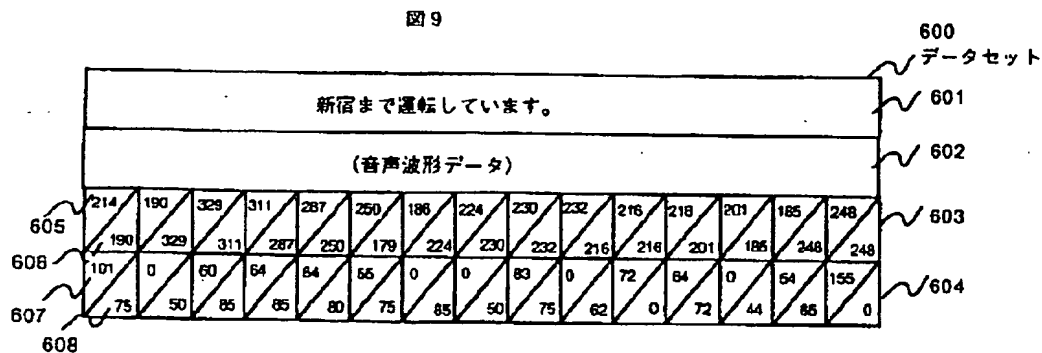


【図 5】

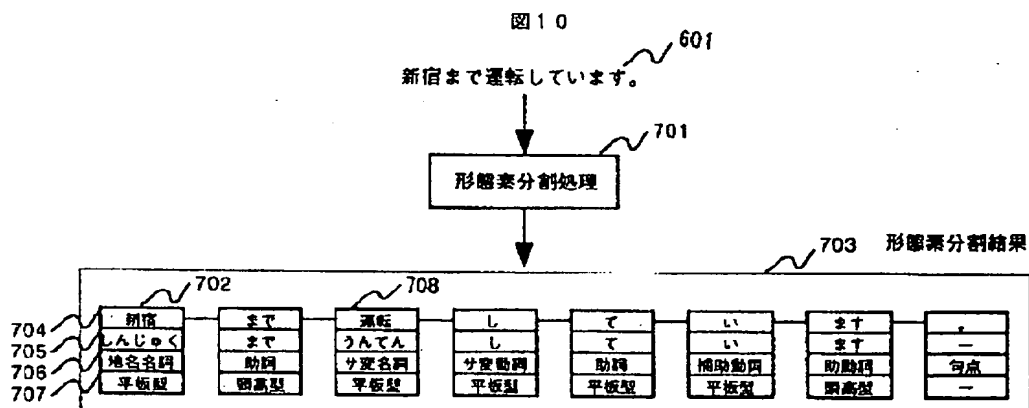
図 5



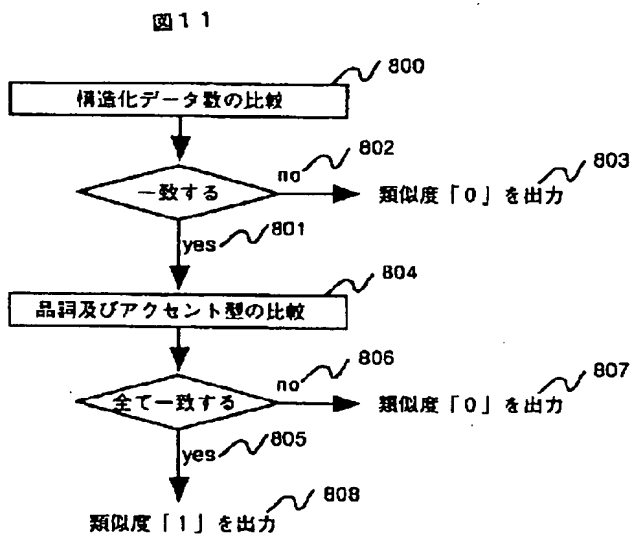
【図9】



【図10】

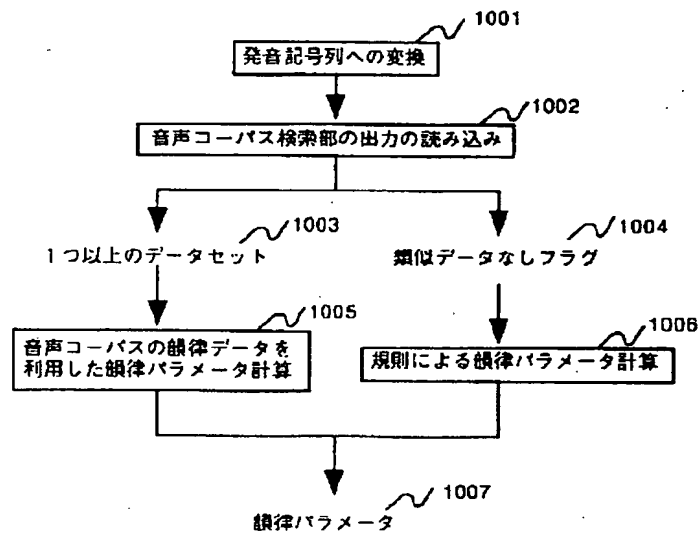


【図11】



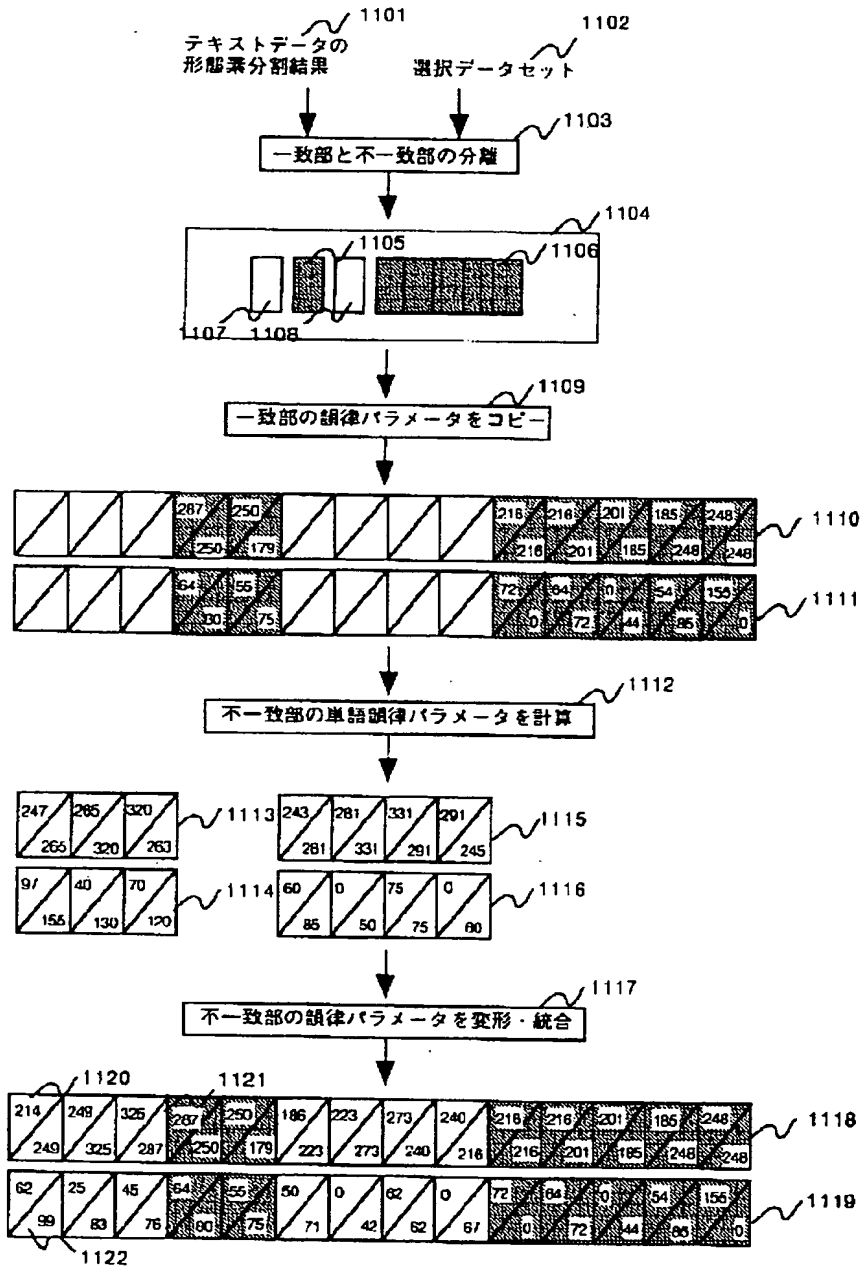
【図 1 2】

図 1 2



【図 14】

図 14



フロントページの続き

(72)発明者 安藤 ハル

東京都国分寺市東恋ヶ窪一丁目280番地  
株式会社日立製作所中央研究所内

(72)発明者 矢島 俊一

東京都小平市上水本町五丁目20番1号 株  
式会社日立製作所半導体事業部内